15

The Management of a Multicamera Tracking System for Videosurveillance by Using an Agent Based Approach

Bethel Atohoun and Cina Motamed
Laboratoire LISIC, Université du Littoral Côte d'Opale, France

1. Introduction

The development of vision systems for monitoring or surveillance of wide area sites is an interesting field of investigation. In order to maximize the capabilities and performance of such system, it is often necessary to use a variety of sensor devices that complement each other. The standard configuration consists in completely covering a scene with a set of cameras with adjacent Fields Of View (FOV). Many people from the computer vision community have worked on the geometrical aspect of this configuration. By using several overlapping calibrated cameras, the system generates a global virtual view of the scene. The use of multiple views of the same scene in the tracking process provides the ability to resolve a part of occlusion situations. A second and less explored configuration is based on a network of non-overlapping cameras. This second configuration is economically attractive because it permits to efficiently decrease the number of sensors. However, the incomplete coverage makes the tracking problem more difficult. The main difficulty is the establishment of correspondence between the objects captured by multiple sensors (cross-camera data association).

In this work we present a high level sensor management strategy in a context of video-surveillance including both of the two configurations: overlapping and distant cameras. The global objective of the system is the development of the object tracking task.

The general problems of multi-sensor management are related to decisions about what sensors to use and for which purposes, as well as when, where and how to use them. This last side of high level management is closely linked with the concept of active perception strategy (Bajcsy, 1998). This strategy is particularly adapted where real-time performance is needed such as tracking, robot navigation, surveillance, visual inspection. The active perception has been widely developed for designing the perception for mobile robotic. In fact real-time perception systems have theirs limitation in the computation of massive amount of input data with processing procedures in a reduced and fixed amount of time. The active strategy has the capacity to filter data and to focus the attention of the perception to relevant information and also can choose the best alternative by using the contextual information. Such approaches are closely linked with the design of cognitive system which permits to combine knowledge and reasoning in order to develop smart and robust perception system.

An important objective of an intelligent multi-sensor system is to exploit the complementarity and the redundancy of sensors. For homogenous sensors the complementarity permits to enlarge the field of perception and the redundancy permit to improve the accuracy of measurements. For a given configuration some area of the scene can be covered by a unique sensor, or by several sensors or none of them. In the case of heterogeneous sensor it permits to increase the number of the perception modalities, in order to cope, for example, with several perception conditions (night/day).

In order to perform the high level management, it is essential to characterize each sensor with respect to its utility for the task. Generally the main characterization of the sensor is given by the measurement uncertainty. Many sensors have a variable uncertainty with respect to their functioning condition. And generally it is difficult to model the uncertainty globally. Another sensor characterization feature concerns its field of coverage.

Tracking is an important task of computer vision with many applications in surveillance, scene monitoring, navigation, sport scene analysis and video database management. One of the most critical parts of a multi-object-tracking algorithm is the data association step. It has to deal with new objects, short or long term disappearing objects and occlusions. Related works linked with the tracking of humans and vehicles with multiple visual sensors are numerous (Snidaro et al., 2004) (Nakazawa et al., 1998) (Utsumi et al., 1998) (Foresti et al, 1995).

Statistical approaches as the kalman filtering approach with its extensions are widely used in tracking and control (Bar-Shalom & Forthmann, 1988). They are particularly adapted for tracking targets in clutter. Qualitative approaches are an alternative for motion correspondence (Veenman et al, 2001)(Sethi & Jain, 1990)(Rangarajan & Shah, 1991). These approaches are less normative than conventional statistical methods used in tracking. The main advantage is their flexibility, because they permit an easy integration of several forms of a priori information and contextual information for constraining complex problem. Our approach belongs globally to the last category.

The section 2 presents the local tracking module. The section 3 details the global strategy of our distributed tracking system. The section 4 represents the management strategy in the presence of overlapping camera FOVs. The section 5 is focused on the problem of tracking based on distant camera without common FOV. Finally, the section 6 shows some experimental results over a real world application.

2. Local tracking module

The goal of this unit is to detect and track object locally in the field of view of each Local Vision System (LVS). The first step concerns the motion detection step. For a fixed sensor, the standard motion detection approach consists in modelling the stationary background which has to be updated in order to tolerate low illumination variation. We have used the Stauffer and Grimson algorithm based on the Mixture Gaussians in order to model the background with multiple possible states (Stauffer & Grimson, 2000). The figure 1 shows an illustration of the moving objects detection.

The local tracking of objects under the FOV of a camera uses a region-based approach. It uses cinematic and visual constraints for establishing correspondence. In our problem of object tracking for a visual-surveillance application, observations representing the detected regions are complex, and are not corrupted by random noise only. They are also affected by detection errors, merging and splitting artefacts, which are difficult to model globally or statistically. In addition, the dynamic model of human activity may present significant

Fig. 1. Moving object segmentation, test images from the dataset PETS'2006, camera C3

variability according to the working context. For all these reasons, we have chosen a qualitative approach, which mainly focuses on the data association quality rather than the precision of the object motion estimation. Assumptions underlying our approach concern several points: objects move smoothly from a frame to another one, objects do not change quickly their appearance colour and objects can interact with other objects or groups of objects. The tracking algorithm uses the Nearest Neighbour (NN) strategy. However, in the presence of merging or splitting situations, two specific procedures are launched in order to solve the association ambiguity.

When a merging situation is detected, a notion of a temporary group of objects is defined in order to track the global region containing visually merged objects. In addition to the temporary group tracking, the algorithm attempts to maintain the track of each individual object inside the global group region. The estimation of position of the individual objects during the merging situation is based on their appearance model by using the Mean Shift algorithm (Comaniciu, 2000) (Cheng, 1995). When the system detects a splitting situation associated with a known group, a specific procedure focuses its attention toward the identity of objects. A visual comparison between objects before merging and after splitting permits to affect the best object identity for each region by using the colour histogram. After splitting, the group updates its individual object. When the group is reduced to a known sole object, the group entity is destroyed [Motamed, 2006].

A planar homography transformation H^S is performed in order to bring each local track estimation from sensor 's' into the same reference plan R_0. The position in image concerns the position of object on the ground plane obtained by the lower segment of the object bounding box. At each image sequence, the output of the local tracker is composed by the position of the tracked object and two complementary information.

3. Global organisation of the distributed multi-camera tracking system

Under a surveillance objective, the system has to react in real time and to manage efficiently its distributed sensors' resources. In particular, in a multi-camera organisation, raw image sequences cannot be easily sent over the network and in addition to this, the central unit has generally not enough capacity to compute alone all information in parallel. The distributed organisation is adapted to reduced network bandwidth constraint. In fact, all low level information can be processed locally and only high-level information is transmitted over the network.

It is important to explicitly manage the uncertainty of the object's identity. In other words, the system has to notify situations where the risk of object confusion becomes important. This capacity of self-evaluation is essential for a safety device and is known as the "positive security".

The proposed system integrates a hybrid strategy including a distributed and central organisation. It is composed of a network of distributed local vision systems (LVS). The system contains a supervisor level, whose role is to collect LVS decisions and performs the global task of high level tracking for surveillance.

In order to manage efficiently the information flow coming from LVSs at the supervisor level, we propose a specific logical organisation at the supervisor level. It is based on a multi-agent society framework, which is considered as an interesting approach when the distributed problem can be represented and solved with a group of cooperative intelligent entities (Matsuyama, 2001). The notion of agent is associated with each tracked object and is defined for each LVS. This organisation permits to join together agents in a specific "society" in order to solve specific tracking task.

As we have mentioned in the introduction the system has to deal with two general categories of multi-camera configuration (Fig. 2). The first is based on overlapping FOV of cameras and the second concerns distant cameras with separate FOVs. In order to efficiently operate with these different configurations, three independent management strategies are jointly implemented:

- The basic strategy is in the case of an object under the FOV of only one sensor. In this situation, the supervisor takes into account directly the result of the local tracker.
- For overlapping configuration, a logical society based on agents from these sensors is created. The society allows the fusion of object position estimations obtained by multiple views.
- For a configuration containing sensors with separate FOVs, another specific logical society is created in order to perform the cross camera object association. The society contains agents that are linked with each neighbouring sensors likely to perceive the object.

The instantaneous result of the local tracking system, encapsulated in the concept of an logical agent, is sent to the supervisor level. The agent contains the current position, indicators and status of the tracked object. Firstly, the identity indicator $Id^t(O_i)$ of each object O_i represents a degree of belief with respect to the identity of the object during the local tracking. The status of the object is associated with the notion of group of objects defined in the local tracking level. If the object belongs to a group, a third indicator delivers the degree of visibility of the object from the sensors. The last important information concerns the fact of entering and exiting the FOV of the local sensor. At each entry or exit the local sensor send to the supervisor a set of information resuming the appearance of object (color histograms, dimension). This last information linked with the entry and exit will be explicitly exploited by the cross camera tracking strategy.

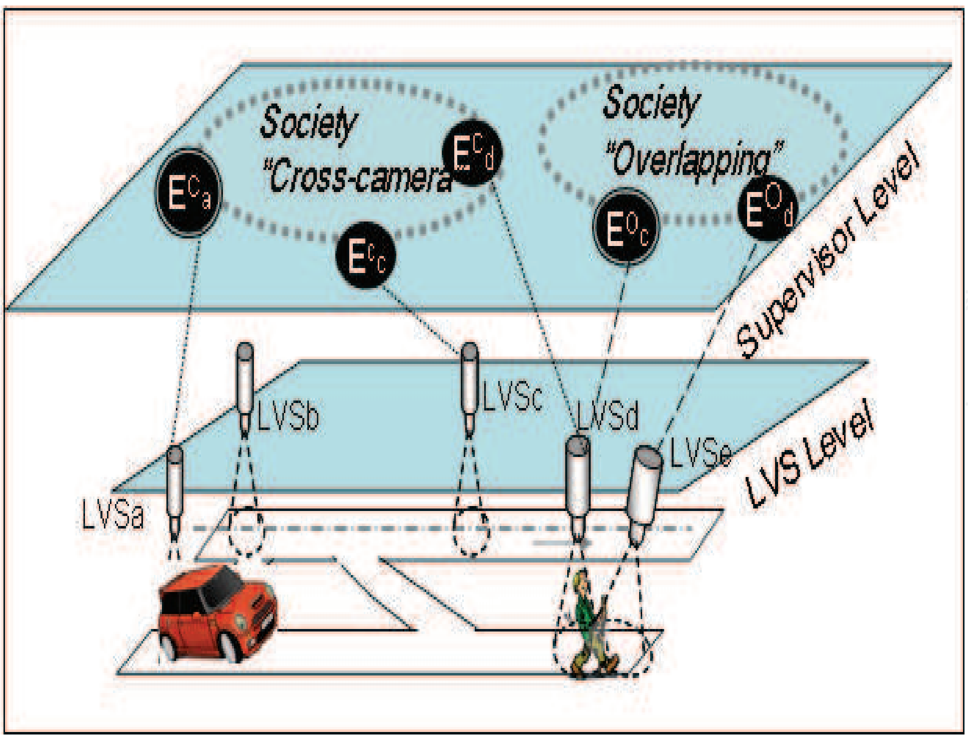


Fig. 2. Organisation of societies for distant cameras configuration (Cross camera) and for overlapping sensor FOVs configuration

4. Tracking with overlapping camera FOVs

In the presence of overlapping FOV, the system has to deal with the limitation of each sensor with respect to the occlusion problem. In the proposed approach, when a tracked object is detected inside an overlapping FOV area a society of agents is automatically created at the supervisor level society for each object. The society contains agents associated with each local tracker. The society operates as a centralized filter by combining agents results (local tracking).

The notion of group and the visibility indicator obtained from the local tracker allows the estimation of the dynamic suitability of a sensor in an overlapping situation with respect to the tracked object. The sensor suitability permits to guide the tracking process by a context driven strategy. The main objective is to favor the best views with respect to dynamic occlusion of each tracked object. This indicator clearly favor un-occluded tracked object with respect to the camera views.

For each object with respect to each sensor 's', the suitability indicator (λ_s) is set to its maximum=1 if the tracked object is known as sole object from the local sensor. When the object goes out of the field of the sensor the indicator takes the zeros value. In presence of occlusion, highlighted by the notion of group managed by the local sensor 's', the indicator takes the degree of visibility estimated during the merging procedure of the local tracker. The visibility degree is obtained directly from the Meanshift estimation in the case of the object occlusions.

For each society, the supervisor level performs a combination of positions estimation from local sensor weighted by their suitability indicators in order to generate a fused position P_{global}.

$$P_{global} = \frac{1}{\sum_s \lambda_s} . \sum_s H^S(P_s^*).\lambda_s \qquad (1)$$

The homography transformation H^S brings each estimation from each sensor 's' into the same reference plan R_0.

5. Tracking with distant cameras

The tracking with distant camera, also known as the cross-camera tracking, can be considered as close to the conventional target tracking (radar or vision), because the trajectories of objects are constructed from observations. The main difference is that in conventional tracking algorithms based on state filtering, observations are obtained synchronously in a continuous space. In cross-camera tracking, objects move between discrete locations and can disappear for a long time (blind zones).

In addition to detection errors, or occlusions at the sensor level, the cross-camera data association over wide scene induces three levels of difficulties:

- There are numerous cases where objects look similar.
- Objects' behaviours between two sensors are unpredictable or coarsely defined.
- Variation of outdoor illumination, sensor response and object orientation, induces changes in the visual appearance of objects.

In order to make the re-identification feasible, in addition to the objects' visual signature, the system has to constrain the object motion in blind zones. It permits to focus the attention on interesting candidates in front of a set of sensors.

(Huang & Russell, 1997) and (Kettnaker & Zabih, 1999) have addressed the problem of data association in cross-camera tracking in the context of the Artificial Intelligence and Computer Vision communities, respectively. Huang's algorithm has been tested in freeway environments, and Kettnaker's algorithm has been designed for human re-identification inside an office. Both algorithms use the Bayesian formalism and attempt to track all possible detected objects over a network of cameras. The data association stage is transformed to a weighted assignment problem solving.

Previous algorithms do not contain an explicit temporal reasoning scheme. However Kettnaker proposes an online extension of their approach by selecting a set of active candidates. The main applications of previous re-identification systems are the link travel time between locations and origin-destination counting, except the kettnaker's system, which is designed for human activity monitoring. Our work is also inspired by the Bayesian approach. It permits to perform the data-association in a distributed manner by integrating prior information about visual and behavioural model of objects.

Once a new object is detected by an LVS, a first level society (Cross camera society) is generated at the supervisor level in order to wait for the future re-appearance of the object over the network. The society contains agents that are linked with each neighbouring LVSs likely to perceive the awaited object. The constitution of the society is performed offline by a human analyst by using contextual knowledge of the scene (topology of the scene and traffic behaviour) and depends on the configuration of the perception system. For example, if an object is detected by the LVS_a, the society, which is in charge of its tracking, will be by default, composed by the agents linked to LVS_a, LVS_c and LVS_d (Fig. 3).

The manager of the society is the agent associated with the LVS that has detected the object. The waiting procedure is activated during a time delay controlled by the temporal constraints. Under this organisation, the association is performed under a strategy of prediction–verification. This network organisation can be seen as an information routing layer from LVS information level to supervisor level. In other terms, each agent focuses on the messages from its associated LVS and waits for its awaited object. The manager sends, to each agent of its society, a message containing visual and temporal constraints associated with the appearance of the awaited object. These messages permit the creation of temporary tracks between the manager and each agent. When an agent E_i re-identifies the awaited object, it sends a message to the manager of the society and confirms the current track of the object. At each re-identification, the manager validates the decision of the association. The existing society is then destroyed, and recursively, the supervisor creates a new one in order to follow the object over the network.

Figure 3 illustrates these main steps during the crossing of a vehicle in front of a distributed network of LVS. At instant T1 the track of the object is initialized and a first society is launched. At instant T2, one of the agents of the current society re-identifies the awaited object and returns this decision to its manager for validation. The first society is then destroyed. The figure at instant T3 shows the second society which has been created to maintain the track of the object.

Temporal constraints are represented in our system by a fuzzy temporal interval called (DOP: Domain Occurrence Possibility) represented by a possibility distribution. The $DOP(O_i,E_m,E_n)$ is the prediction generated by the agent E_m, explaining the temporal appearance possibility of an awaited object O_i in the field of a specific agent E_n. Each society manager has to generate these coarse temporal constraints for each agent of its society. The DOP between two nodes is estimated by the kinematic equation:

$$d = v_0.t + \frac{A.t^2}{2}$$

$$=> \; t = \frac{-v_0 + \sqrt{{v_0}^2 + 2.A.d}}{A} \tag{2}$$

$$or \quad t = \frac{d}{v_0} \, , \qquad (if\, A=0)$$

with

V_0 : the measured local speed of the object

A : the model of object acceleration between the two nodes

d : the distance d between the two nodes

The computation of the prediction 't' has to integrate the acceleration variability and also the measurement errors. For this, all variables are approximated by a normal distribution with respect to their mean and variance values. The distribution of the variable $t(m_t,\sigma_t)$ is computed by using specific arithmetic operators for the normal distributions (Courtney & Thacker, 2001). This procedure allows the propagation of the theoretical errors over the expression of t. The DOP is then built by using the normal distribution of the variable t. A standard transformation from a normal distribution to a trapezoidal possibilistic distribution is applied. The trapezoidal model is centred at the instant m_t with a core and a base of 2. σ_t and 3. σ_t respectively. A human analyst coarsely initialises generic acceleration models A_i by using the contextual knowledge for each class of object Ci (human, car etc…). They represent essentially physical limitations of the class behaviour between two nodes.

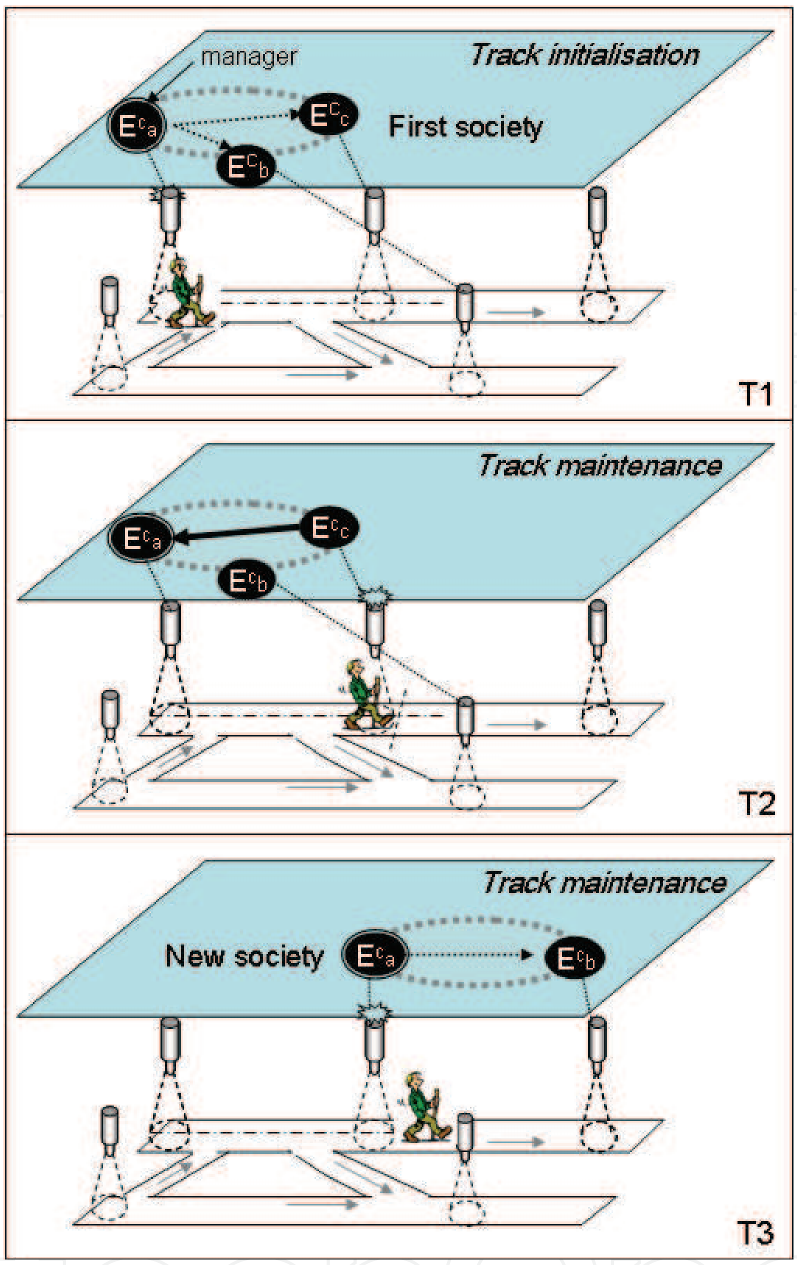


Fig. 3. Illustration of the creation of the first level "cross camera" society

In a complex multi-sensor network, there is no effective one to one correspondence between awaited objects and observed objects. A standard instantaneous approach has some limitations with asynchronous observations. In fact, in some ambiguous situations, the system may have not received all interesting information. An efficient approach is the "temporal fusion" strategy. This last strategy tries to improve dynamically the quality of decisions with more observations. In fact, for a distributed multi-sensor architecture with no common field of view, the system has to exploit the complementarities of observations over time and over the fields of sensors. The global compatibility between two objects is obtained by the product of their visual compatibility with their temporal compatibility. The temporal compatibility value represents the value of the DOP at the instant of the object detection. The proposed distributed tracking approach uses as in conventional tracking systems three main steps linked to track initialisation, maintenance, and termination.

The strategy of creation of new objects is chosen close to the Open-world assumption existing in the Demspter Shafer Theory of Evidence. In other terms, when the observation is not compatible with any of the awaited objects, the occurrence of the new object is favoured. The track termination of an object is decided when an awaited object is not detected by its current society during its lifespan controlled by its DOPs.

When several objects are tracked simultaneously, multiple societies have to work together. For each observation, within a LVS, measures of compatibility with all awaited objects represent a local distribution of preference. The validation of an association H_i by an agent is performed by computing the possibility $P(H_i)$ and the necessity $N(H_i)$ of the association. The possibility distribution is then built by normalizing the distribution of preference inside the segment [0,1]. The necessity of a hypothesis translates the notion of uniqueness of the hypothesis with respect to other alternatives. If the necessity is high, it means that no ambiguity is present. In this situation, the agent can validate the association solely. When the necessity of the association is low, a notion of ambiguity is declared.

$$N(H_i) = 1 - P(\bar{H}_i)$$

With
$$\bar{H}_i = \Phi - \{H_i\} \tag{3}$$

$$P(\bar{H}_i) = \max_{H_j \in \bar{H}_i} P(H_j)$$

We present in figure 4, the two situations of ambiguity appearing during the association of two visually compatible objects. Dop1 and Dop2 represent predictions of object O1 and object O2, respectively, on the downstream sensor. Obs1 and Obs2 are the observations of O1 and O2 detected by the downstream sensor. In the first scenario (Type 1), the first observation, with respect to the predictions cannot be associated to a unique object. The second observation can remove the ambiguity by using a temporal fusion strategy. In the second scenario (Type 2), the two observations are detected within the intersection of the two predictions. The system decides that it cannot make the association. However, the system reacts quickly, once it has detected the two observations, without waiting for the end of awaited objects lifespan.

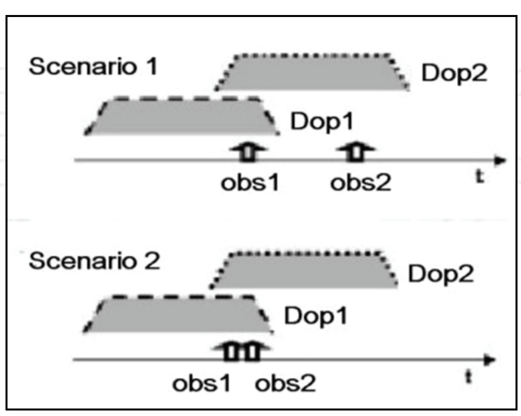


Fig. 4. Scenarios of ambiguity Type 1 and Type 2 with two similar objects

When a first level agent detects an ambiguity containing two or more objects, a second level society is generated by integrating all concerned first level societies containing ambiguous objects. The objective of the second level society is to develop a Temporal Fusion procedure. In

order to compare efficiently all hypotheses, the temporal fusion uses a MHT tree (Multiple Hypotheses Testing) (Reid, 1979). When a second level society is activated, the attached first level societies stop their own decisions and wait for the decisions of the temporal fusion.
The MHT tree describes all possible hypotheses resulting from observations obtained by the set of LVS concerned by the second level society. A MHT hypothesis integrates a set of associations with their global degree of compatibility. For each MHT hypothesis, a possibility and a necessity measurement is estimated. The possibility measurement of a MHT hypothesis is built by the product of the compatibility of its associations. At each observation, the tree is updated. A pruning procedure removes a MHT hypothesis when its possibility is low (< 0.1). The data association decision is then deferred as long as the confidence level of one of MHT hypothesis is not significant enough compared to the other ones. This decision is controlled by the necessity of the hypotheses. The MHT tree is stopped when the relative necessity measurement of a MHT hypothesis is sufficient (> 0.3).
Figure 5 illustrates a basic example where the temporal fusion strategy efficiency resolves an ambiguity associated to a scenario of type 1 in a configuration with four LVS. The ambiguity concerns two visually similar vehicles (Object_1, Object_2) appearing at instants T1 et T2.

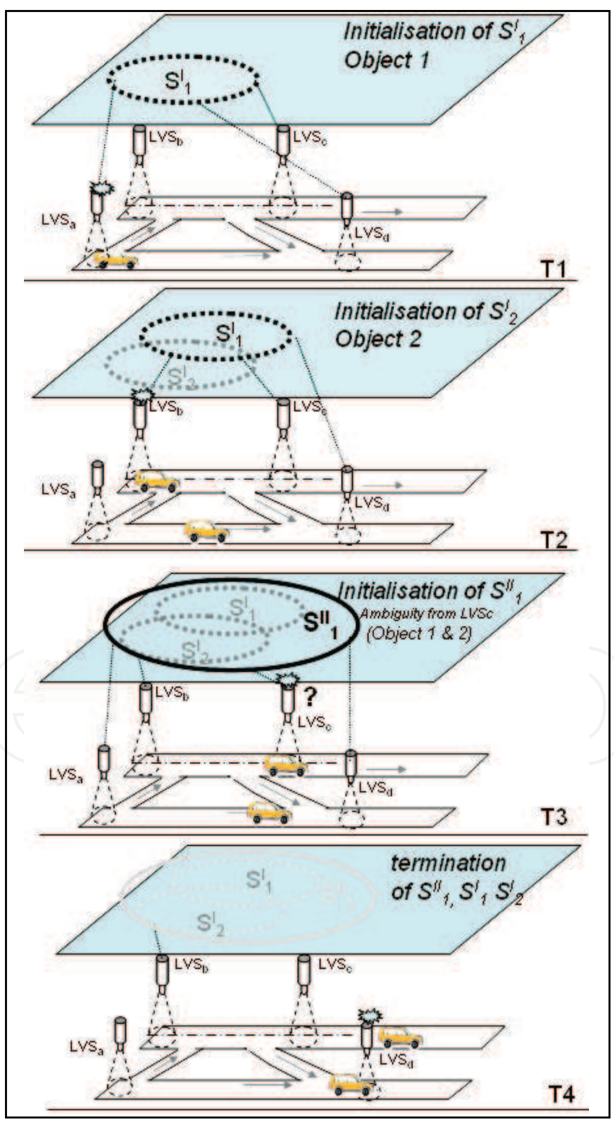


Fig. 5. Resolution of scenario of type 1

The detection of the Object_1 at the instant T1 from LVS_a, induces the creation of a first level society S^I_1 to perform the tracking of the object. At instant T2, another society S^I_2 is created for the Object_2. The observation of the an object Obs(T3) (at instant T3) on the LVS_c generates an ambiguity for experts of S^I_1 and S^I_2, linked with LVS_c. In fact, two awaited objects are visually similar and temporally compatible with the observation. So a second level society S^{II}_1 is created at instant T3, in order to perform a temporal fusion strategy. The figure 6 shows the predictions (DOPs) used by experts linked to LVS_c and LVS_d.

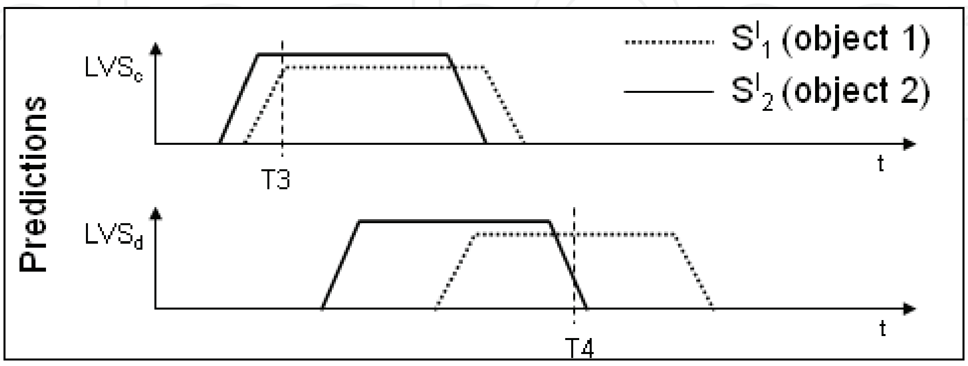


Fig. 6. Temporal prediction for the example

In our example, a second observation Obs(T4) from LVS_d permits the elimination of the ambiguity (table 1). The first observation generates two hypotheses 1.1 and 1.2 indicating a low necessity. Hypothesis 1.1 represents the association of the Obs(T3) to objet_1. Hypothesis 1.2 represents the association of the obs(T3) to object_2. After the second observation Obs(T4), Hypothesis 2.1 with a necessity value of (0.73) is accepted and finally the Obs(T3) is associated to Object_1 and the Obs(T4) to Object_2.

MHT Hyp. N°	Obs(T3) From LVSc (compatibility degree)	Obs(T4) From LVSd (compatibility degree)	Possibility of the MHT hyp.	Necessity of the MHT hyp.
1.1	(1.0) /Object_1		**1.00**	**0.10**
1.2	(0.9) /Object_2		**0.90**	**0.00**
2.1	(1.0) /Object _1	(1.0) /Object_2	**1.00**	**0.73**
2.2	(0.9) /Object_2	(0.3) /Object_1	**0.27**	**0.00**

Table 1. a MHT tree with the degree of possibility and necessity of the hypotheses

Unfortunately, there are three distinct cases where the Temporal Fusion decides that it cannot resolve the ambiguity:

- The first case is when all observations are detected and the MHT tree is performed without success. It means that the system has detected a scenario of ambiguity of type 2 (fig.4). This situation is detected by the MHT tree, when the necessity of each hypothesis remains low. The temporal fusion is then stopped after the last awaited observation.
- The second case is when a part of awaited objects is not detected at the end of predicted DOPs. The temporal fusion is stopped at the end of their DOPs. These problems may occur when objects do not verify their prediction, and also when objects are not correctly detected by the LVSs.

- The third case is when the number of awaited objects, interacting within the MHT tree, becomes too important (object count > 5). The temporal fusion prematurely stops in order to avoid the explosion of the MHT tree.

Decisions of success or non-decision obtained by the temporal fusion are sent to managers of each primary society. Then, in their turn, they forward their final decisions to the supervisor. In a surveillance context, as we have mentioned in the section 3, this notion of a controlled non-decision is more acceptable than standard errors.

6. Experiments

In this section we present a first experiment based on the proposed global management approach. Real sequences are obtained by recording synchronised image sequences from four cameras observing a campus. The figure 7 illustrates the multi-camera configuration of the scene.

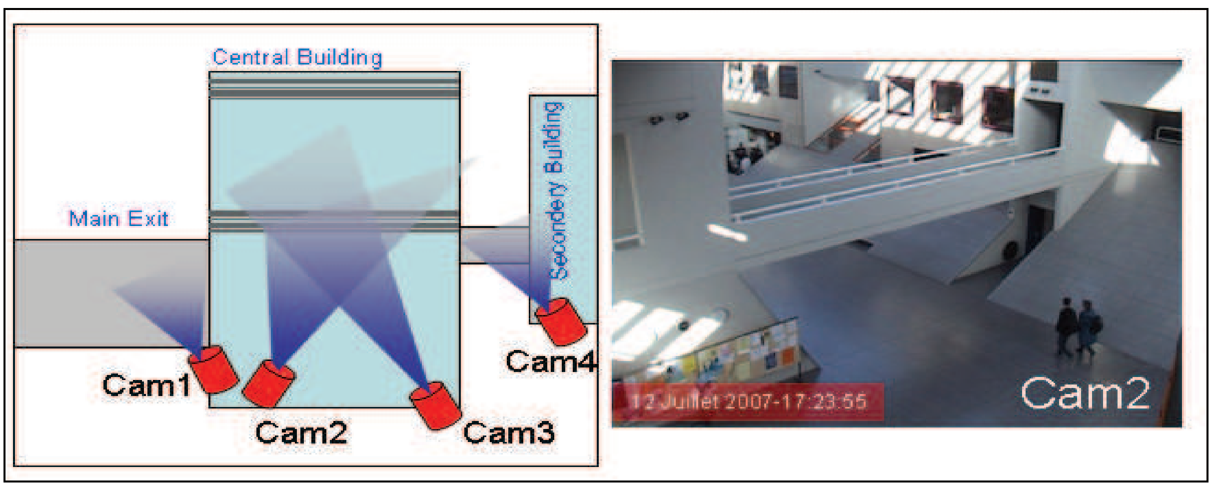


Fig. 7. The multi-camera configuration of the observed campus and an image from the second camera.

The table 2 summarizes the results of the data association decisions with a comparison with straightforward (basic) techniques. In fact, for cross camera tracking the basic NN algorithm decides the association of the observation with the best instantaneous candidate, and in the case of overlapping configuration, the track maintenance is performed only by the local tracker who has primary detected the object. The evaluation of performance of our system with respect to the above basic techniques is obtained by using ground truth data delivered by a human analyst.

The third line represents the notion of ambiguity rate (non decision rate) of our system which is essentially due to the resemblance of a part of tracked objects from the distant cameras societies. In the context of security such non-decision score is more acceptable than standard errors. In this experiment the global tracking results are satisfactory and the sum of the errors and non decision rate of the proposed approach is lower than the global tracking errors of the basic algorithm.

When the count of simultaneous compatible objects (temporally and visually) is low, the cross camera association based on the TF algorithm reacts efficiently with an acceptable non-decision rate. Otherwise the performance of the TF algorithm decreases significantly. In fact, an excessive augmentation of object, firstly favours scenarios of ambiguity of type 2 (fig. 4) which are detected but are not solved by the temporal fusion and secondly, it induces an excessive number of awaited objects interacting in the MHT trees. In this last situation, the temporal fusion prematurely stops in order to avoid the explosion of the MHT trees size.

Results	Proposed management strategy	Basic approach : NN algorithm (in distant configuration) & without track fusion (in overlapping situation)
Global count of objects (ground truth)	122	122
Correct global track (in term of identity)	90	69
non decision (distant camera)	23	X
Global Track errors	9	53

Table 2. Illustration of performance and comparison with basic algorithms

7. Conclusion

We have proposed an agent-based architecture in context of a distributed vision based tracking system. The objective of the system is the tracking of objects over a wide area scene by using a high level multi-sensor management strategy.
The main originality of this work with previous works concerns the capacity of management of multi-camera systems for surveillance including both overlapping cameras FOV and distant cameras configuration. The proposed system has been successfully tested in a real indoor environment.
Our strategy of sensor management naturally copes explicitly with known modelled tracking ambiguities. At the local level the tracking takes into account the regions merging, splitting, and target missing. In the presence of muti-view configuration of the scene, it can solve efficiently a good part of occlusions situation. And finally, in the presence of distant cameras, the strategy of temporal fusion allows to control temporally the tracking system decisions over the time.
Future works concern the improvement of the survivability of such distributed tracking system with respect to the loss of LVSs by performing an automatic reconfiguration of the sensor network. It will permit to generate active societies based only on operational LVSs. This reconfiguration may be controlled as in conventional computer network by using a kind of “keep-alive” messages between LVS and the supervisor.

8. References

Bajcsy, R. (1998). Active Perception, *Proceedings of the IEEE*, 76 (8) (1998) 996-1005.
Bar-Shalom, Y. & Forthmann, T.E. (1988). *Tracking and data association*. Academic Press.
Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol 17(8), 1995, pp.790-799.
Comaniciu, D.; Ramesh, V. & Meer P. (2000). Real-TimeTracking of Non-Rigid Objects using Mean Shift. *IEEE Computer Vision and Pattern Recognition,* Vol II, 2000, pp.142-149.

Courtney, P.; Thacker, N.A. (2001). Performance characterization in computer vision: The role of statistics in testing and design. *In Imaging and Vision Systems : Theory, Assessment and Applications, Jacques Blanc-Talon and Dan Popescu (Eds), 2001,* NOVA Science Book.

Foresti G.L.; Murino, V.; Regazzoni, C. & Vernazza, G. (1995). A multilevel fusion approach to object identification in outdoor road scenes. *International Journal of Pattern Recognition and Artificial Intelligence*, 1995, 9(1):23–65.

Huang, T. & Russell, S. (1997). Object identification in a Bayesian context. *In International Joint Conference on Artificial Intelligence* (1997), 1276-1282.

Kettnaker, V. & Zabih, R. (1999). Bayesian multi-camera surveillance. *In IEEE Conference on Computer Vision and Pattern Recognition (*1999) 253-259.

Matsuyama, T. (2001) Cooperative distributed vision: dynamic integration of visual perception, camera action, and network communication. *4th international Workshop on Cooperative Distributed Vision* (2001), Kyoto Japan.

Motamed, C. (2006). Motion detection and tracking using belief indicators for an automatic visual-surveillance system. *Image and Vision Computing*, Volume 24, Issue 11, 1, Pages 1192-1201, 2006.

Nakazawa, A.; Kato, H. & Inokuchi, S. (1998). Human tracking using distributed vision systems. *In Proceedings of the 14th ICPR*, pages 593–596.

Rangarajan, K. & Shah, M. (1991). Establishing Motion Correspondence. CVGIP: Image Understanding, 54 (1), (1991) pp. 56-73.

Reid, B. (1979). An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, AC-24(6) (1979) 843-854.

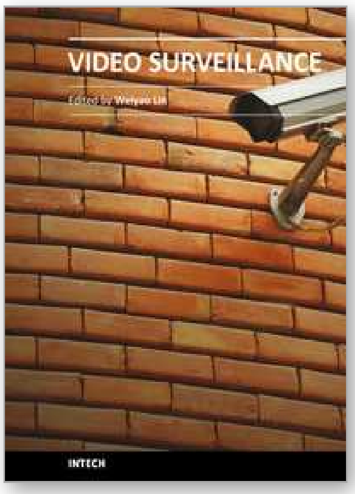
Sethi, I.K. & Jain, R. (1990). Finding trajectories of feature points in a monocular image sequence. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 9 (1), 56-73.

Snidaro, L.; Luca Foresti, G.; Niu, R. & Varshney, P. K. (2004). Sensor fusion for video surveillance, International Conference on Information Fusion (2004),pp. 739-746, Sweden.

Stauffer, C. & Grimson, W. (2000). Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8) (2000) 747–757.

Utsumi, A.; Mori, H.; Ohya, J. & Yachida, M. (1998). Multiple view-based tracking of multiple humans. *In Proceedings of the 14th ICPR*, pages 597–601.

Veenman, C.J.; Reinders, M.J.T.; & Backer, E. (2001). Resolving Motion Correspondence for Densely Moving Points. *IEEE Transactions on PAMI* vol23, pp. 54-72.



Video Surveillance
Edited by Prof. Weiyao Lin

ISBN 978-953-307-436-8
Hard cover, 486 pages
Publisher InTech
Published online 03, February, 2011
Published in print edition February, 2011

This book presents the latest achievements and developments in the field of video surveillance. The chapters selected for this book comprise a cross-section of topics that reflect a variety of perspectives and disciplinary backgrounds. Besides the introduction of new achievements in video surveillance, this book also presents some good overviews of the state-of-the-art technologies as well as some interesting advanced topics related to video surveillance. Summing up the wide range of issues presented in the book, it can be addressed to a quite broad audience, including both academic researchers and practitioners in halls of industries interested in scheduling theory and its applications. I believe this book can provide a clear picture of the current research status in the area of video surveillance and can also encourage the development of new achievements in this field.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821